

Shape Expression Schema における Property Path を用いた問合せの充足可能性判定

前田 祐紀

近年, RDF データ(グラフデータ)の普及が進んでいる. また, グラフデータのサイズは非常に大きいのが一般的である. グラフデータに対して問合せを行う際, ユーザはデータの構造を十分に把握しているとは限らないため, 充足不能な問合せが記述されることがある. ここで, どの妥当なグラフデータに対しても解が空である問合せを充足不能であるという. 充足不能な問合せの実行は明らかに無駄であり, 回避することが望ましい. そこで本研究では, グラフデータにおける問合せ式の充足可能性判定を行う手法について考える.

本研究では, スキーマ言語として Shape Expression Schema (ShEx), 問合せ言語として Property Path を用いる. RDF データのスキーマ言語として RDF Schema が提案されているが, オントロジー定義言語としての性質が強く, スキーマ定義には必ずしも向いていないため, 本研究では ShEx を対象とする. また, 本研究の対象とする Property Path は, SPARQL 1.1 の一部として定義されている問合せ言語である. 先行研究として, ShEx に対するパターン問合せの充足可能性判定を行う手法が提案されているが, Property Path では開始状態から終了状態までのルートが固定されないため, パターン問合せに比べてより柔軟な探索が可能となっている.

提案手法は, オートマトンに基づいて構成されている. まず, Property Path をオートマトン M_P に変換する. 次に, ShEx をオートマトン M_G に変換する. そして M_G の各ノード n を M_P の開始状態とした時に M_P が終了状態まで到達できるかを調べる. 具体的には, M_P と M_G での現在いるノードから同じラベルで遷移できるノードを調べていき, M_P が終了状態まで遷移できるかを調べる. 到達できるルートがあれば充足可能, 一つもなければ充足不能と判定できる.

提案手法を Ruby により実装し, 評価実験を行った. まず, SP²Bench を用いて RDF データを生成した. ここで, SP²Bench は DBLP に基づき任意のトリプル数もしくはデータサイズの RDF データを生成する SPARQL ベンチマークである. なお, SP²Bench には ShEx が付与されていないため, ShEx は著者が作成した. そして, 充足不能な Property Path を作成し, RDF データに対する問合せと ShEx に対する充足可能性判定に要する実行時間を計測した. その結果, 後者の前者に対する比率は非常に小さく, 問合せ実行に対する充足可能性判定のオーバーヘッドは極めて小さいことが分かった.

(指導教員 鈴木伸崇)